A Novel Extended Granger Causal Model Approach Demonstrates Brain Hemispheric Differences During Face Recognition Learning

Tian Ge[†]   Keith M. Kendrick[*]   Jianfeng Feng[†,‡,**]

[†] Centre for Computational Systems Biology, Fudan University, PR China

[*]Cognitive and Systems Neuroscience Group, The Babraham Institute, Cambridge CB22 3AT, UK

[‡]Centre for Scientific Computing, University of Warwick, Coventry CV4 7AL, UK

[**] Corresponding author

[**] E-mail: jianfeng64@gmail.com

October 1, 2009

Abstract

Two main approaches in exploring causal relationships in biological systems using time-series data are the application of Dynamic Causal model (DCM) and Granger Causal model (GCM). These have been extensively applied to brain imaging data and are also readily applicable to a wide range of temporal changes involving genes, proteins or metabolic pathways. However, these two approaches have always been considered to be radically different from each other and therefore used independently. Here we present a novel approach which is an extension of Granger Causal model and also shares the features of the bilinear approximation of Dynamic Causal model. We have first tested the efficacy of the extended GCM by applying it extensively in toy models in both time and frequency domains and then applied it to local field potential recording data collected from *in vivo* multi-electrode array experiments. We demonstrate face discrimination learning-induced changes in inter- and intra-hemispheric connectivity and in the hemispheric predominance of theta and gamma frequency oscillations in sheep inferotemporal cortex. The results provide the first evidence for connectivity changes between and within left and right ITs as a result of face recognition learning.

Author Summary

Causality analysis has become one of the most powerful and valuable tools in discovering connections and building networks in complex biological systems. Here, we developed a novel causal model to extend Granger Causal model and also share the features of the bilinear approximation of Dynamic Causal model which are two prominent techniques in causality analysis. We applied our extended model to local field potential data recorded in the left and right temporal cortices of sheep during face recognition learning and have provided the first evidence for inter and intra-hemispheric connectivity changes in both time and frequency domains following learning. We also observed significant asymmetry between the two hemispheres which suggested that learning is more associated with global encoding and low frequency oscillations in the right hemisphere while the left hemisphere is more involved in a local encoding strategy which predominates after learning.

1 Introduction

In order to exploit the full potential of high-throughput data in biology we have to be able to convert them into the most appropriate framework for contributing to knowledge about how the biological system generating them is functioning. This process is best conceptualized as first building a nodal network derived from empirically derived knowledge of the biological structures and molecules involved (nodes) and then secondly to use computational-based steps to discover the nature, dynamics and directionality of connections (directed edges) between them.

Causality analysis based upon experimental data has become one of the most powerful and valuable tools in discovering connections between different elements in complex biological systems [1–6]. However, despite some encouraging successes in various areas in systems or computational biology its development and application have been impeded by a number of issues about the meaning of causality. For example, in clinical science, the current emphasis on how to apply causality approaches mainly resides in resolving the problem of how clearly to define causality itself [7]. A typical problem cited is the so called "Simpson paradox" in which the successes of groups seem reversed when the groups are combined. This demonstrates the ambiguity that can result in determining causal relationships based only on frequency data. However, this issue disappears if we incorporate time into the definition of causality as Granger has done in the field of Economics [8]. Nevertheless there is still no accepted unified way to tackle this issue. Taking altered gene expression data using microarray analysis as an instance, there are three approaches one can use to deal with the time-series data obtained: the simple dynamical system approach, the dynamical

Bayesian network approach and the Granger causality approach which is a generalization of the dynamical system one. In [9], we have discussed in detail the pros and cons of applying the latter two approaches and shown potential advantages in using Granger when sufficient repeated measurements are available. With brain activity data from functional magnetic resonance imaging (fMRI) experiments, two prominent techniques have been introduced to address temporal dependencies and directed causal influences: Dynamic Causal (DCM) and Granger Causal (GCM) models. These two models have always been considered to differ radically from each other [10, 11]. DCM establishes state variables in the observed data and is believed to be a causal model in a true sense. On the other hand, GCM is a phenomenological model which just tests statistical dependencies among the observations to determine how the data may be caused [10–12]. The importance of the two approaches in interpreting fMRI data is demonstrated in [10] and in 2008 there were around 450 papers published devoted to both approaches and excluding those relating to other types of biological data.

The key question we want to address in the current paper is whether we can develop an extended and biophysical constraint approach to share the features of the various approaches mentioned above, and in particular of the two causal models: DCM and GCM? The significance of such an approach is obvious and we would expect that its application could represent a powerful new tool in systems and computational biology, particularly in association with increasingly powerful genomic, proteomic and metabolic methodologies allowing time-series measurements of large numbers of putatively interacting molecules.

In this paper, we will show that GCM can be extended to a more biophysical constraint model by incorporating some features of the bilinear approximation of DCM. By setting up

5

a conventional VAR model with additional deterministic inputs and observation variables, we can create a more general model: Extended Granger Causal Model (EGCM) which offers a new way to establish connectivity.

The EGCM is first tested in two toy models. With both state and observation variables, the interactions between nodes are successfully recovered using an extended Kalman filter approach and partial Granger to establish causality in both time (DCM) and frequency (GCM) domains respectively. The GCM approach itself is not tailored particularly well for biological experiments where we are often faced with the case of the data being recorded with and without a stimulus present. The time gap between two adjacent stimuli is very short and we would expect the network structure to remain unchanged during the whole experiment although the form and the intensity of the input may be unknown. This scenario is also the case for the gene network data considered in [1,2], where the authors have treated the two situations separately, although there should be a common and true structure for both. We have therefore also used EGCM in toy models to establish its efficacy in revealing the true network structure when there is an intermittent input to affect state variables.

To exemplify the direct application of EGCM to establishing causality in a specific biological system, we have applied it to local field potential (LFP) data recorded in the sheep inferotemporal cortex (IT) of both left and right hemispheres before and after they learn a visual face discrimination task [13]. There is electrophysiological, molecular neuroanatomical and behavioral evidence for asymmetrical processing of faces in the sheep brain similar to humans [14–16] although cells in both the left and right IT respond selectively to faces [15]. Learning also alters both local and population based encoding in sheep IT as well as theta-nested gamma frequency oscillations in both hemispheres and there is

6

greater synchronization of theta across electrodes in the right IT than there is in the left IT [13]. There is considerable interest in establishing functional differences between the ways the left and right brain hemispheres interact and process information [17–19]. It has recently been hypothesized that the left hemisphere specializes in controlling routine and tends to focus on local aspects of the stimulus while the right hemisphere specializes in responding to unexpected stimuli and tends to deal with the global environment [18, 19]. Establishing altered causal connections and frequency dependency within and between the two hemispheres during face recognition learning will help test this hypothesis.

2 Methods

2.1 EGCM Model

The traditional and widely used Granger Causal Model takes the form [20]:

$$\vec{x}(t) = A_1 \vec{x}(t-1) + \cdots + A_p \, \vec{x}(t-p) + \vec{\varepsilon}(t) \tag{1}$$

where $\vec{x}(t) \in \mathbb{R}^N$, $A_i = \begin{bmatrix} a_{11}^i & \cdots & a_{1N}^i \\ \vdots & \ddots & \vdots \\ a_{N1}^i & \cdots & a_{NN}^i \end{bmatrix}$, $i = 1, \cdots, p$ are coefficient matrices, $\vec{\varepsilon}(t)$ is the noise, and the model has a vector autoregressive representation with an order up to p.

In spite of its successful application, GCM requires the direct observation of the state variables and does not include designed experimental effects in the model which form some of its limitations. Here we extend GCM to a more reasonable and biophysical constraint model by incorporating additional deterministic inputs and observation variables, closely

following equations which are the features in the Dynamical Causal Model and its bilinear approximation form [10]. The extended Granger Causal Model takes the form:

$$
\begin{cases}
\vec{x}(t) &= [A_1 + u(t-1)B_1]\vec{x}(t-1) + \cdots \\
&\quad + [A_p + u(t-p)B_p]\vec{x}(t-p) + v(t-1)\vec{c} + \vec{\varepsilon}(t) \\
\vec{y}(t) &= \tilde{g}(\vec{x}(t)) + \vec{\varepsilon^*}(t)
\end{cases} \tag{2}
$$

where $u(t)$ and $v(t)$ are deterministic inputs, $\vec{y}(t)$ are the observation variables which are the function $\tilde{g}$ of the state variables, $B_i, i = 1, \cdots, p$ are the coefficients that allow the inputs to modulate the coupling of the state variables, $\vec{\varepsilon}(t)$ and $\vec{\varepsilon^*}(t)$ are intrinsic and observation noise and are mutually independent.

Now, if we can recover the state variables $\vec{x}(t)$ from the noise observation variables $\vec{y}(t)$, all the problems can be considered in the framework of the traditional Granger causality. It's clear that the formal difference between EGCM and GCM is that we have included observation variables and deterministic inputs which are assumed to be known and will affect the connection of the state variables as well as the state variables directly. However, EGCM also has a strong connection with Dynamical Causal Model. We refer the readers to Discussion section for a detailed discussion on the importance of this extension, in particular the relationship between Eq. (2) and the Volterra type series expansion [10].

2.2 EGCM Algorithm

For the extended Granger Causal model Eq. (2), we now introduce an algorithm to estimate the state variables as well as all its parameters which will give us the first inspiration of the connection of the state variables.

$$\text{Let } \vec{X}(t) = \begin{bmatrix} \vec{x}(t) \\ \vdots \\ \vec{x}(t-p+1) \end{bmatrix}, \vec{Y}(t) = \begin{bmatrix} \vec{y}(t) \\ \vdots \\ \vec{y}(t-p+1) \end{bmatrix}, \vec{U}(t) = \begin{bmatrix} u(t) \\ \vdots \\ u(t-p+1) \end{bmatrix}.$$

Then, the VAR(p) model can be reduced to a VAR(1) model which takes the form:

$$\vec{X}(t+1) = \begin{bmatrix} A_1 + u(t)B_1 & A_2 + u(t-1)B_2 & \dots & A_p + u(t-p+1)B_p \\ I & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} \vec{X}(t)$$

$$+ v(t) \begin{bmatrix} \vec{c} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \vec{w}(t)$$

$$\triangleq A(\vec{\theta}, \vec{U}(t))\vec{X}(t) + C(\vec{\theta})v(t) + \vec{w}(t) \triangleq \vec{f}(\vec{X}(t), \vec{U}(t), v(t), \vec{\theta}) + \vec{w}(t),$$

$$\vec{Y}(t) = \vec{h}(\vec{X}(t)) + \vec{s}(t)$$

where $\vec{\theta}$ is the parameter vector to be estimated. $\vec{w}(t)$ and $\vec{s}(t)$ are both zero-mean uncorrelated Gaussian noise with covariance matrix $Q(t)$ and $R(t)$ respectively.

In order to apply the model to real data, we have to estimate both the states and parameters of the model from input variables and noise observations. A widely used method for this dual estimation is extended Kalman filter (EKF) [21, 22]. Here we recursively approximate the nonlinear system by a linear model and use the traditional Kalman filter for the linearized model.

Let $\vec{\xi} = \left[\vec{X}^T, \vec{\theta}^T\right]^T$ then

$$
\begin{aligned}
\vec{\xi}(t+1) &= \begin{bmatrix} \vec{X}(t+1) \\ \\ \vec{\theta}(t+1) \end{bmatrix} = \begin{bmatrix} \vec{f}(\vec{X}(t), \vec{U}(t), v(t), \vec{\theta}) \\ \\ \vec{\theta}(t) \end{bmatrix} + \begin{bmatrix} \vec{w}(t) \\ \\ \vec{\eta}(t) \end{bmatrix} \\
&= \begin{bmatrix} A(\vec{\theta}, \vec{U}(t))\vec{X}(t) + C(\vec{\theta})v(t) \\ \\ \vec{\theta}(t) \end{bmatrix} + \vec{\zeta}(t) \\
&\triangleq \vec{g}(\vec{\xi}(t), \vec{U}(t), v(t)) + \vec{\zeta}(t)
\end{aligned}
$$

where $\vec{\eta}(t)$ is uncorrelated Gaussian noise with covariance matrix $Z(t)$. Define

$$
\begin{aligned}
\hat{\xi}_{t|t} &= \mathbb{E}[\xi(t)|\vec{Y}(t), \vec{U}(t), v(t)] \\
\hat{\xi}_{t+1|t} &= \mathbb{E}[\xi(t+1)|\vec{Y}(t), \vec{U}(t), v(t)] \\
\Omega_{t|t} &= \mathbb{E}[(\xi(t) - \hat{\xi}_{t|t})(\xi(t) - \hat{\xi}_{t|t})^T|\vec{Y}(t), \vec{U}(t), v(t)] \\
\Omega_{t+1|t} &= \mathbb{E}[(\xi(t+1) - \hat{\xi}_{t+1|t})(\xi(t+1) - \hat{\xi}_{t+1|t})^T|\vec{Y}(t), \vec{U}(t), v(t)]
\end{aligned}
$$

where $\hat{\xi}_{t|t} = \begin{bmatrix} \hat{X}_{t|t} \\ \\ \hat{\theta}_{t|t} \end{bmatrix}, \hat{\xi}_{t+1|t} = \begin{bmatrix} \hat{X}_{t+1|t} \\ \\ \hat{\theta}_{t+1|t} \end{bmatrix}$. Then, the EKF algorithm for dual estimation

consists of two steps: prediction and updating.

Prediction: Given the estimated state $\hat{\xi}_{t|t}$, the observation $\vec{Y}(t)$ and inputs $\vec{U}(t)$ and $v(t)$,

we predict the state variables and the covariance matrix of prediction error of the system at

time $t+1$.

$$\hat{\xi}_{t+1|t} = \mathbb{E}[\xi(t+1)|\vec{Y}(t), \vec{U}(t), v(t)] = \mathbb{E}[\vec{g}(\xi(t), \vec{U}(t), v(t)) + \vec{\zeta}(t)|\vec{Y}(t), \vec{U}(t), v(t)]$$

$$\approx \mathbb{E}[(\vec{g}(\hat{\xi}_{t|t}, \vec{U}(t), v(t)) + \frac{\partial \vec{g}}{\partial \xi^T}(\xi(t) - \hat{\xi}_{t|t})|\vec{Y}(t), \vec{U}(t), v(t)]$$

$$= \vec{g}(\hat{\xi}_{t|t}, \vec{U}(t), v(t)) = \begin{bmatrix} A(\hat{\theta}_{t|t}, \vec{U}(t)) & 0 \\ 0 & I \end{bmatrix} \hat{\xi}_{t|t} + \begin{bmatrix} C(\hat{\theta}_{t|t}) \\ 0 \end{bmatrix} v(t)$$

$$\Omega_{t+1|t} = \mathbb{E}[(\xi(t+1) - \hat{\xi}_{t+1|t})(\xi(t+1) - \hat{\xi}_{t+1|t})^T|\vec{Y}(t), \vec{U}(t), v(t)]$$

$$= \mathbb{E}[(\vec{g}(\xi(t), \vec{U}(t), v(t)) + \vec{\zeta}(t) - \hat{\xi}_{t+1|t})(\vec{g}(\xi(t), \vec{U}(t), v(t))$$

$$+ \vec{\zeta}(t) - \hat{\xi}_{t+1|t})^T|\vec{Y}(t), \vec{U}(t), v(t)]$$

$$\approx \mathbb{E}[(\frac{\partial \vec{g}}{\partial \xi^T}(\xi(t) - \hat{\xi}_{t|t}) + \zeta(\vec{t}))(\frac{\partial \vec{g}}{\partial \xi^T}(\xi(t) - \hat{\xi}_{t|t}) + \zeta(\vec{t}))^T|\vec{Y}(t), \vec{U}(t), v(t)]$$

$$= F_t \Omega_{t|t} F_t^T + \Psi(t)$$

where $\Psi(t) = \begin{bmatrix} Q(t) & 0 \\ 0 & Z(t) \end{bmatrix}$ and

$$F_t = \begin{bmatrix} \dfrac{\partial f}{\partial X^T} & \dfrac{\partial f}{\partial \theta^T} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A(\vec{\theta}, \vec{U}(t)) & \dfrac{\partial}{\partial \theta^T}[A(\vec{\theta}, \vec{U}(t))\vec{X} + C(\vec{\theta})v(t)] \\ 0 & I \end{bmatrix}_{\theta = \hat{\theta}_{t|t}, X = \hat{X}_{t|t}}$$

Updating: We use the new observation $\vec{Y}(t+1)$ at time $t+1$ to update the state of the system.

$$\hat{\xi}_{t+1|t+1} = \hat{\xi}_{t+1|t} + G(t+1)[\vec{Y}(t+1) - \vec{h}(\hat{X}_{t+1|t})]$$

$$\Omega_{t+1|t+1} = [I - G(t+1)H(t+1)]\Omega_{t+1|t}$$

where

$$G(t+1) = \Omega_{t+1|t}H^T(t+1)[H(t+1)\Omega_{t+1|t}H^T(t+1) + R(t+1)]^{-1}$$

$$H(t+1) = \begin{bmatrix} \dfrac{\partial h}{\partial X^T} & 0 \end{bmatrix}_{\hat{X}_{t+1|t}}$$

2.3 EGCM Definition of Causality

After recovering the state variables using the EGCM algorithm above, we can define the causality with the idea proposed by Granger. The only difference is that, in our EGCM model, two deterministic inputs $u(t)$ and $v(t)$ are added to the normal autoregressive representation. Here, we provide the formulation of EGCM causality in both time domain and frequency domains.

2.3.1 Causality in the Time Domain

For simplicity of notation, here we only formulate EGCM for two time series X_t and Y_t. To generalize them to more general case of multi time series, we refer the readers to [23, 24]. Assume that X_t and Y_t in our EGCM have the following representation:

$$
\begin{aligned}
X_t &= \sum_{j=1}^{p}[a_{1j} + u(t-j)b_{1j}]X_{t-j} + C_{1x}v(t-1) + \epsilon_{1t} \\
Y_t &= \sum_{j=1}^{p}[d_{1j} + u(t-j)e_{1j}]Y_{t-j} + C_{1y}v(t-1) + \epsilon_{2t}
\end{aligned}
\tag{3}
$$

A joint representation in our EGCM that includes the past information of both processes X_t and Y_t can be written as:

$$
\begin{aligned}
X_t &= \sum_{j=1}^{p}[a_{2j} + u(t-j)b_{2j}]X_{t-j} + \sum_{j=1}^{p}[d_{2j} + u(t-j)e_{2j}]Y_{t-j} + C_{2x}v(t-1) + \epsilon_{3t} \\
Y_t &= \sum_{j=1}^{p}[f_{2j} + u(t-j)g_{2j}]X_{t-j} + \sum_{j=1}^{p}[h_{2j} + u(t-j)k_{2j}]Y_{t-j} + C_{2y}v(t-1) + \epsilon_{4t}
\end{aligned}
\tag{4}
$$

where p is the maximum number of lagged observations included in the model. $\epsilon_{it}(t)$, $i = 1, 2, 3, 4$ are prediction errors with variance Σ_i and are uncorrelated over time. Then, according to the causality definition of Granger, if the prediction of one process can be improved by incorporating the past information of the second process, then the second pro-

cess causes the first process. So, in the extended model here, we define that if the variance of prediction error for the process X_t is reduced by the inclusion of the past information of the process Y_t, then, a causal relation from Y_t to X_t exists. This can be quantified as

$$F_{Y \to X} = \ln \frac{\Sigma_1}{\Sigma_3} \tag{5}$$

If $F_{Y \to X} = 0$, there is no causal influence from Y_t to X_t and if $F_{Y \to X} > 0$, there is. Similarly, we can define the causal influence from X_t to Y_t as

$$F_{X \to Y} = \ln \frac{\Sigma_2}{\Sigma_4} \tag{6}$$

2.3.2 Causality in the Frequency Domain

Our EGCM also allows a frequency domain decomposition to detect the intrinsic causal influence which provides valuable information.

We define the lag operator L to be $LX_t = X_{t-1}$ and assume here that the input $u(t)$ is a constant, i.e. $u(t) \equiv u$ to avoid the appearance of nonlinearity. Then, the joint representation of both processes X_t and Y_t in equation (4) can be expressed as:

$$
\begin{aligned}
X_t &= \sum_{j=1}^{p}[a_{2j} + ub_{2j}]X_{t-j} + \sum_{j=1}^{p}[d_{2j} + ue_{2j}]Y_{t-j} + C_{2x}v(t-1) + \epsilon_{3t} \\
&= \sum_{j=1}^{p}\tilde{a}_{2j}X_{t-j} + \sum_{j=1}^{p}\tilde{b}_{2j}Y_{t-j} + C_{2x}v(t-1) + \epsilon_{3t} \\
Y_t &= \sum_{j=1}^{p}[f_{2j} + ug_{2j}]X_{t-j} + \sum_{j=1}^{p}[h_{2j} + uk_{2j}]Y_{t-j} + C_{2y}v(t-1) + \epsilon_{4t} \\
&= \sum_{j=1}^{p}\tilde{c}_{2j}X_{t-j} + \sum_{j=1}^{p}\tilde{d}_{2j}Y_{t-j} + C_{2y}v(t-1) + \epsilon_{4t}
\end{aligned} \tag{7}
$$

Rewrite equation (7) in terms of lag operator, we have:

$$
\begin{bmatrix} \tilde{a}_2(L) & \tilde{b}_2(L) \\ \tilde{c}_2(L) & \tilde{d}_2(L) \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \epsilon_{3t} \\ \epsilon_{4t} \end{bmatrix} + v(t-1) \begin{bmatrix} C_{2x} \\ C_{2y} \end{bmatrix} \tag{8}
$$

13

where $\tilde{a}_2(0) = 1$, $\tilde{b}_2(0) = 0$, $\tilde{c}_2(0) = 0$, $\tilde{d}_2(0) = 1$.

Since what we really care about is the causal relationship caused by the intrinsic connection of the state variables rather than the outside driving force, i.e. the input $v(t)$, after fitting the model (7) and getting the covariance matrix of the prediction error, we just go on with the following model:

$$
\begin{bmatrix} \tilde{a}_2(L) & \tilde{b}_2(L) \\ \tilde{c}_2(L) & \tilde{d}_2(L) \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \epsilon_{3t} \\ \epsilon_{4t} \end{bmatrix} \tag{9}
$$

which means that after fitting the EGCM with input $v(t)$ to eliminate outside influence, we just focus on the intrinsic causal influence in the frequency domain.

After normalizing equation (9) using the transformation proposed by Geweke [25, 26] to eliminate the cross term in the spectra, we assume that we have the normalized equation in the form:

$$
\begin{bmatrix} \bar{a}_2(L) & \bar{b}_2(L) \\ \bar{c}_2(L) & \bar{d}_2(L) \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \bar{\epsilon}_{3t} \\ \bar{\epsilon}_{4t} \end{bmatrix} \tag{10}
$$

Fourier transforming both sides of equation (10) leads to

$$
\begin{bmatrix} \bar{a}_2(\omega) & \bar{b}_2(\omega) \\ \bar{c}_2(\omega) & \bar{d}_2(\omega) \end{bmatrix} \begin{bmatrix} X(\omega) \\ Y(\omega) \end{bmatrix} = \begin{bmatrix} \bar{E}_x(\omega) \\ \bar{E}_y(\omega) \end{bmatrix} \tag{11}
$$

Recasting equation (11) into the transfer function format we obtain

$$
\begin{bmatrix} X(\omega) \\ Y(\omega) \end{bmatrix} = \begin{bmatrix} H_{xx}(\omega) & H_{xy}(\omega) \\ H_{yx}(\omega) & H_{yy}(\omega) \end{bmatrix} \begin{bmatrix} \bar{E}_x(\omega) \\ \bar{E}_y(\omega) \end{bmatrix} \tag{12}
$$

After proper ensemble averaging we have the spectral matrix

$$
S(\omega) = H(\omega) \Sigma H^*(\omega) = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} \tag{13}
$$

14

where * denotes the complex conjugate and matrix transpose and $\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$ is

the covariance matrix of the prediction errors in equation (11). Hence, we can define the

causal influence from Y_t to X_t at frequency ω as

$$f_{Y \to X}(\omega) = \ln \frac{S_{xx}(\omega)}{H_{xx}(\omega)\Sigma_{xx}H_{xx}^*(\omega)} \tag{14}$$

Similarly, we can define the causal influence from X_t to Y_t at frequency ω as well.

Note that although here we just provide the definition of pairwise Granger causality

for EGCM, it's obvious that similar methods can be easily applied to the definition of

conditional, partial or complex Granger causality in both time and frequency domains [9,

23, 24, 27–31]. Since the explicit meaning of the parameters in the EGCM (i.e. the intrinsic

coupling among state variables, the strength of the inputs to modulate the coupling and

the influence of the inputs on the state variables directly), we can also get an idea of the

connection of the state variables and how the inputs affect them from the fitted model before

we translate it into a single number.

2.4   Methods for LFP Recording Experiment

Animals and visual discrimination training. Three female sheep were used (Ovis aries,

one Clun Forest and two Dorsets). All experiments were performed in strict accordance

with the UK 1986 Animals Scientific Procedures Act. During the experiments the animals

were housed inside in individual pens. They were trained initially over several months

to perform operant-based face (sheep) or non-face (objects) discrimination tasks with the

animals making a choice between two simultaneously presented pictures, one of which was

associated with a food reward. During stimulus presentations, animals stood in a holding trolley and indicated their choice of picture by pressing one of two touch panels located in the front of the trolley with their nose. The food reward was delivered automatically to a hopper between the two panels. The life-sized pictures were back projected onto a screen 0.5m in front of the animal using a computer data projector. A white fixation spot on a black background was presented constantly in between trials to maintain attention and experimenters waited until the animals viewed this spot before triggering presentation of the image pairs. The stimulus images remained in view until the animal made an operant response (generally around 1-3 s). In each case, successful learning of a face or object pair required that a performance criterion of >75% correct choices over 40 trials (i.e. 40 pairs) was achieved consistently. By the end of training, animals were normally able to reach the > 75% correct criterion after 40-80 trials and maintain this performance. For the current analysis extensive recordings taken during and after learning of novel face pairs were used (2 pairs for Sheep A; 2 pairs for Sheep B – only one of which was successfully learned – and 1 pair for Sheep C). In all cases recordings were made over 20-80 trials during learning and then during 46-170 trials after the > 75% correct criterion was reached. Post learning trials ranged from within 5-10 minutes of the end of a learning trial session to 2 months after learning. For the face pairs, Sheep A and B were discriminating between the faces of different unfamiliar sheep faces (face identity discrimination) whereas for Sheep C, discrimination was between calm and stressed face expressions in the same animal (face emotion discrimination). With this latter animal, the calm face was the rewarded stimulus.

Electrophysiological recordings and analysis. Following initial behavioral training sheep were implanted under general anesthesia (fluothane) and full aseptic conditions with unilat-

eral (Sheep A-right IT) or bilateral (Sheep B and C) planar 64-electrode arrays (epoxylite coated, etched, tungsten wires with 250 μm spacing - total array area 2mm x 2mm tip diameter < 1 m, electrode impedence 200-300kΩ) aimed at the IT. Holes (0.5cm diameter) were trephined in the skull and the dura beneath cut and reflected. The electrode bundles were introduced to a depth of 20-22 mm from the brain surface using a stereotaxic micro-manipulator and fixed in place with dental acrylic and stainless-steel screws attached to the skull. Two of these screws acted as reference electrodes, one for each array. Electrode depths and placements were calculated with reference to X-rays, as previously described. Electrodes were connected to 34 pin female plugs (2 per array) which were cemented in place on top of the skull (using dental acrylic). Starting 3 weeks later, the electrodes were connected via male plugs and ribbon cables to a 128 channel electrophysiological recording system (Cerebus 128 Data Acquisition System - Cyberkinetics Neurotechnology Systems Inc., USA) and recordings made during performance of the different face and non-face pair operant discrimination tasks. This system allowed simultaneous recordings of both neuronal spike and local event-related (LFP) activity from each electrode. Typically, individual recording sessions lasted around 30 min and for 80-200 individual trials. There was at least a week between individual recording sessions in each animal. The LFPs were sampled at 2kHz and digitized for storage from around 3 seconds prior to the stimulus onset to around 3 seconds after the stimulus onset (stimulus durations were generally 1-3 s).

For data analysis of the stored signals LFP data contaminated with noise such as from animal chewing food were excluded as were LFPs with unexpectedly high power. For LFPs, offline filtering was applied in the range of 1-200 Hz and trend was removed before spectral analysis. Any trial having more than 5 points outside the mean ± 5 standard devi-

17

ation range were discarded before analysis. At the end of the experiments, animals were euthanized with an intravenous injection of sodium pentobarbitone and the brains removed for subsequent histological confirmation of X-rays that array placements were within the IT cortex region.

3 Results

In order to evaluate the performance of EGCM for the estimation of the state variables as well as the prediction of the parameters, we first applied the method to two toy models.

3.1 Toy Models

3.1.1 Toy Model 1

The first toy model we used comes from a traditional VAR model which has been extensively applied in tests of Granger causality [32]. We modified the model by adding two deterministic inputs $u(t)$ and $v(t)$. $u(t)$ was assumed to be a constant stimulation, i.e. $u(t) \equiv 0.1$ while $v(t)$ was assumed to a harmonic oscillator and had the form of a sinusoidal function since biological rhythms are a common phenomenon [27]. Observation variables were also included in the toy model and assumed to be nonlinear functions of the state variables since it's a real challenge to uncover state variables with nonlinear mapping from states to measurements [33, 34]. We generated the time series according to the

following equations:

$$
\begin{aligned}
x_1(t) &= 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + 0.1x_1(t-1) \\
&\quad + 0.1\cos[\tfrac{2\pi}{50}(t-1)] + \varepsilon_1(t) \\
x_2(t) &= -0.5x_1(t-1) + 0.5x_3(t-2) - 0.8 \times 0.1x_2(t-1) + \varepsilon_2(t) \\
x_3(t) &= -0.5x_2(t-1) + 0.5x_3(t-1) + 1.2 \times 0.1x_3(t-1) \\
&\quad - 0.1\cos[\tfrac{2\pi}{50}(t-1)] + \varepsilon_3(t)
\end{aligned}
$$

where $\varepsilon_i(t)$, $i = 1, 2, 3$ were zero mean uncorrelated Gaussian noise with variance 0.5, 0.8 and 0.6 respectively. Hence, according to the general form of EGCM, in this toy model we have:

$$
A_1 = \begin{bmatrix} a_{11}^1 & a_{12}^1 & a_{13}^1 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 \\ a_{31}^1 & a_{32}^1 & a_{33}^1 \end{bmatrix} = \begin{bmatrix} 0.95\sqrt{2} & 0 & 0 \\ -0.5 & 0 & 0 \\ 0 & -0.5 & 0.5 \end{bmatrix},
$$

$$
A_2 = \begin{bmatrix} a_{11}^2 & a_{12}^2 & a_{13}^2 \\ a_{21}^2 & a_{22}^2 & a_{23}^2 \\ a_{31}^2 & a_{32}^2 & a_{33}^2 \end{bmatrix} = \begin{bmatrix} -0.9025 & 0 & 0 \\ 0 & 0 & 0.5 \\ 0 & 0 & 0 \end{bmatrix},
$$

$$
B_1 = \begin{bmatrix} b_{11}^1 & b_{12}^1 & b_{13}^1 \\ b_{21}^1 & b_{22}^1 & b_{23}^1 \\ b_{31}^1 & b_{32}^1 & b_{33}^1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.8 & 0 \\ 0 & 0 & 1.2 \end{bmatrix}, \quad B_2 = 0, \quad \vec{c} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},
$$

$$
\tilde{A}_1 = \begin{bmatrix} \tilde{a}_{11}^1 & \tilde{a}_{12}^1 & \tilde{a}_{13}^1 \\ \tilde{a}_{21}^1 & \tilde{a}_{22}^1 & \tilde{a}_{23}^1 \\ \tilde{a}_{31}^1 & \tilde{a}_{32}^1 & \tilde{a}_{33}^1 \end{bmatrix} = A_1 + uB_1 = \begin{bmatrix} 1.4435 & 0 & 0 \\ -0.5 & -0.08 & 0 \\ 0 & -0.5 & 0.62 \end{bmatrix},
$$

$$
u(t) \equiv 0.1, \quad v(t) = 0.1\cos[\tfrac{2\pi}{50}(t-1)].
$$

Inspection of the above equations reveals that $x_1(t)$ is a direct source to $x_2(t)$, $x_2(t)$ and $x_3(t)$ share a feedback loop. There is no direct connection between the remaining pairs of the state variables.

Fig. 1 around here

Fig. 1A is an example of the 2000 time-steps of the data and Fig. 1B shows the network structure.

The observation variables were

$$
\begin{aligned}
y_1(t) &= x_1(t) + \varepsilon_4(t) \\
y_2(t) &= x_2(t) + x_3(t) + \varepsilon_5(t) \\
y_3(t) &= x_1(t) \cdot x_3(t) + \varepsilon_6(t)
\end{aligned}
$$

where $\varepsilon_i(t)$, $i = 4, 5, 6$ were zero mean uncorrelated Gaussian noise with variance 0.1 and also uncorrelated with $\varepsilon_i(t)$, $i = 1, 2, 3$.

Now, we can apply the method to this toy model, i.e., to estimate all the parameters $\tilde{A}_1, A_2, \vec{c}$ and state variables $x_i(t)$ from the deterministic input $u(t)$, $v(t)$ and noise observations. Simulations were performed for 2 seconds (2000 equally spaced time points). Fig. 1C shows that the parameters converged to their true values with only small fluctuations after several hundred data points, even though no prior knowledge was included and the initial values of the parameters were assigned to zeros. It has already been pointed out that the covariance matrix $Z(t)$ (see Methods section) of the noise in the parameter equation will affect the convergence rate and tracking performance [35]. In the situation here, a steep decay of the covariance matrix will lead to a better accuracy but the convergence is then slow. On the other hand, a slow decay will lead to a faster convergence but a larger fluctuation is observed. Hence, $Z(t)$ was carefully controlled to reduce to zero as the t

increased (see Fig. 1C).

After the state variables being recovered, we computed the partial Granger causality in both time and frequency domains (see Fig. 1D) and used the bootstrap approach to construct confidence intervals. Specifically, we simulated the fitted model to generate a data set of 1000 realizations of 2000 time points and use 3σ as the confidence interval. In this result, a causal connection was illustrated as part of the network if, and only if, the lower bound of the 95% confidence interval of the causality was greater than zero. The results show that our extended model can detect the causal relationship correctly in both time and frequency domains.

3.1.2 Toy model 2

When dealing with real data it is quite common that we need to detect the causal influence between time series from several variables affected by some stimulus. The stimulus may be very complicated, or hard to measure, and it may be impossible to formulate its form explicitly. However, if we ignore the influence of these inputs and use a traditional VAR model to detect the causality it is quite probable that we will get a misleading structure even if we use a high-order VAR model.

We used the following toy model which has exactly the same connection coefficients between the three state variables considered in Toy model 1 with an additional simple constant input function p:

$$
\begin{aligned}
x_1(t) &= 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + p + \varepsilon_1(t) \\
x_2(t) &= -0.5x_1(t-1) + 0.5x_3(t-2) + p + \varepsilon_2(t) \\
x_3(t) &= -0.5x_2(t-1) + 0.5x_3(t-1) - p + \varepsilon_3(t)
\end{aligned}
\tag{15}
$$

where $\varepsilon_i(t)$, $i = 1, 2, 3$ were zero mean uncorrelated Gaussian noise with variance 0.5, 0.8 and 0.6 respectively.

Here, we assumed that $p \equiv 0.5$ and the observation variables $y_i(t)$, $i = 1, 2, 3$ were identical to the state variables with observation noise. The variance of the noise was 0.1. It is obvious that the network structure is the same one as shown in Fig. 1B. However, if we ignore the constant input and just use a VAR model to detect this structure, we obtain the structure shown in Fig. 2A and Fig. 2B with confidence intervals where two additional causal relationships (i.e. $1 \rightarrow 3$ and $3 \rightarrow 1$) are presented showing that the real causal influence can no longer be correctly detected. Furthermore, when the input is not taken into consideration, the coefficients of the connection matrix will be meaningless and no longer provide us with the correct estimation of the strength of connection strengths between the state variables. This illustrates why we consider it necessary to incorporate the stimulus into our model although sometimes we don't know its form or intensity.

With our extended model we can, to some extent, solve the above issue and detect the causal influence correctly amongst those state variables affected by some unknown stimulus intermittently, although our model is originally set up for deterministic inputs.

We next generated a time series of 10000 time points which was composed of 10 segments with equal length, a.e., $1 - 1000, 1001 - 2000, \cdots, 9001 - 10000$. Each segment took the form:

$$
\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = A_1 \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \\ x_3(t-1) \end{bmatrix} + A_2 \begin{bmatrix} x_1(t-2) \\ x_2(t-2) \\ x_3(t-2) \end{bmatrix} + p \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \end{bmatrix} \tag{16}
$$

where $A_1 = \begin{bmatrix} 0.95\sqrt{2} & 0 & 0 \\ -0.5 & 0 & 0 \\ 0 & -0.5 & 0.5 \end{bmatrix}$, $A_2 = \begin{bmatrix} -0.9025 & 0 & 0 \\ 0 & 0 & 0.5 \\ 0 & 0 & 0 \end{bmatrix}$ and $\varepsilon_i(t)$, $i = 1, 2, 3$

were zero mean uncorrelated Gaussian noise with variance 0.5, 0.8 and 0.6 respectively.

The five segments $1 - 1000, 2001 - 3000, \cdots, 8001 - 9000$ were generated according to the above toy model without input, i.e., $p \equiv 0$, while the remaining five segments were assumed to include input of random intensity which would also affect the state variables randomly. Specifically, within each segment, p was assigned a random value which was generated with the normal distribution $p \sim N(0, 1)$, and the same was the case with c_i: $c_i \sim N(0, 1)$, $i = 1, 2, 3$. Observation variables were still assumed to be identical to the state variables with the variation of observation as 0.1.

Hence, the network structure of the three state variables is still the same as shown in Fig. 1B while each state variable is affected by some input that we don't know the intensity of. Fig. 2C and Fig. 2D show the predicted network structure with confidence intervals using our extended model. The results show that we can still detect the causal influence correctly in this situation.

Fig. 2 around here

3.2 LFP from Left and Right Hemisphere

Local field potential data were obtained from 64-channel multielectrode arrays implanted in the right and left inferior temporal cortices of three sheep (one sheep only had electrodes in the right hemisphere) as previously described [13]. Recordings were made while the animals were presented with pairs of faces which they were required to discriminate be-

tween using an operant response in order to obtain a food reward. In between face pair presentations the animals were presented with a visual fixation stimulus (a white spot on a black screen). Recordings were made during sessions of 20-40 trials where they were either still learning the discrimination or had successfully achieved the learning criterion ($> 75\%$ correct choice of rewarded face). In both the left and right IT the main oscillatory frequencies recorded are in the theta (4-8Hz) and gamma (30-70Hz) ranges and these two frequencies are coupled (theta phase and gamma amplitude) [13]. We have previously shown that learning increases theta amplitude, the ratio of theta to gamma, theta/gamma coherence and the tightness of theta phase [13].

With these experimental data, we can directly use our EGCM to detect the global network for all electrodes in both brain hemispheres. However, due to the size of the network, there are at least a few thousand free parameters to fit. To avoid this issue, we adopt another approach here. For each session we randomly select 3 time series from each region respectively and apply our model to detect the network structure for the six electrodes. This procedure is repeated for 100 times for each session (see Fig. 3 for such an example). The visual stimulus to the IT (including feedforward and feedback signals) is impossible to know in the experiment. However, as we have shown above, we are able to make the assumption that the effect of the stimulus can be regarded as a constant input to each electrode. The inclusion of the stimulus signal will certainly make the model more reasonable.

A further problem here is that if we intend to reconstruct the connections for each six electrodes (left and right) before and after the stimulus respectively, we could end up with two different structures for the time series (not shown). This is certainly not the case since the duration of the stimuli is quite short (1-3 seconds) and the connections will not change

24

in such a short time. To recover a reasonable structure of the connection in these areas in the brain, we therefore assume here that the connections in each trial don't change and the time series before and after the stimulus are generated from a unified structure. With the application of our EGCM approach, we can include the intermittent stimulus and obtain a comparatively reliable structure. Fig. 3 (top-panel) shows such an example where three electrodes in the left and three in the right are randomly selected and that inter- and intra-hemisphere interactions are detected. Fig. 3 (bottom-panel) is the corresponding frequency decomposition of the top panel.

Fig. 3 around here

Fig. 4 around here

In Fig. 4 we show the mean connections within and between the left and right IT calculated using EGCM and as a function of learning. The results clearly demonstrate an asymmetry between the hemispheres. The top-panel is an illustration of the bottom panel which summarizes the results of all experimental data for the two sheep. The most noticeable change is a decrease in the number of connections from the left to the right and an increase in connections within the right but not in the left IT. Indeed there was a strong negative correlation between the number of left to right connections and the number within the right IT for both animals (Sheep B, $r = -0.7739$ ($p=0.0114$); Sheep C, $r = -0.965$ ($p=0.0346$)). These changes occurred as soon as the learning criterion was successfully achieved in successive blocks of trials (i.e. in as little as 5-10 minutes in the case sheep B where learning was successfully achieved within a specific recording session) and were maintained after 1 month or more post-learning. They were not simply the results of stimulus repetition because in Sheep B where recordings were made in repeated sessions of up to 120 trials

but where the learning criterion was not achieved for one of the face pairs there were no connectivity changes observed.

One of the advantages of the extended approach is that we have a frequency domain decomposition. Brain rhythms, not surprisingly, have also been intensively investigated in the literature [36]. Here we concentrate on the two main frequency bands: theta band (4-8 Hz) and gamma band (30-70 Hz) present in our IT recording data and which have been extensively linked to mechanisms of learning and memory [13, 36]. Using the frequency decomposition of our extended model discussed in Methods section, we looked at the following two quantities:

$$\text{mean ratio} \quad = \quad \text{mean interaction in the theta band / mean interaction in the gamma band}$$

$$\text{max ratio} \quad = \quad \text{max interaction in the theta band / max interaction in the gamma band}$$

Fig. 5A shows the mean and maximum ratio integrating the data from all the three sheep in the experiment and at different stages of learning. From this it can be seen that both the mean and maximum ratios in the right hemisphere IT are about double those in the left hemisphere. This clearly indicates that for the right hemisphere, the theta band interaction is more dominant, i.e., the right hemisphere deals more with signals of lower frequency.

Fig. 5 around here

In order to provide a deeper insight into this frequency story, we compute the ratios at different stages of learning. Fig. 5B shows the mean and maximum ratio in two sheep before learning, after learning and a month after learning. An additional set of data is one week after learning for sheep C only. The most noticeable change is the reduction in the interactions in the theta band (low frequencies) in the right IT which occurs after learning

and is maintained subsequently. Combining Fig. 4C (right) with Fig. 5C, we see that learning in general changes the connections in the right hemisphere (increasing), however, the increasing interactions are mainly due to the enhancement of the interaction at the high frequency domain.

4   Discussion

Comparing EGCM, GCM and DCM. EGCM has a strong connection with DCM as well as GCM. We consider the Dynamical Causal model:

$$\begin{cases} d\vec{x}(t) & = & f(\vec{x}, u, \theta)dt + \sigma dB_t \\ \\ \vec{y}(t) & = & \tilde{g}(\vec{x}(t)) + \vec{\varepsilon}(t) \end{cases} \tag{17}$$

where $\vec{x}(t) = (x_1, \cdots, x_N)^T$ are state variables and $(\cdot)^T$ is the transpose of a vector, $u(t)$ is a known deterministic input corresponding to designed experimental effects, θ is the set of parameters to estimate, σ is the diffusion matrix (could depend on time) and B_t is the Brownian motion (or in general, it could be a martingale). The state variables $\vec{x}(t)$ enter a specific model to produce the outputs $\vec{y}(t)$ with the observation noise $\vec{\varepsilon}(t)$.

Here we focus on the bilinear approximation of the Dynamical Causal model which is the most parsimonious but useful form [10]:

$$\begin{cases} d\vec{x}(t) & = & [A + u(t)B]\,\vec{x}(t)dt + u(t)\vec{c}dt + \sigma dB_t \\ \\ \vec{y}(t) & = & \tilde{g}(\vec{x}(t)) + \vec{\varepsilon}(t) \end{cases} \tag{18}$$

where

$$A = \frac{\partial f}{\partial x} = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix}, B = \frac{\partial^2 f}{\partial x \partial u} = \begin{bmatrix} b_{11} & \cdots & b_{1N} \\ \vdots & \ddots & \vdots \\ b_{N1} & \cdots & b_{NN} \end{bmatrix}, \vec{c} = \frac{\partial f}{\partial u} = \begin{bmatrix} c_1 \\ \vdots \\ c_N \end{bmatrix}$$

are parameters that mediate the intrinsic coupling among states, allow the inputs to modulate the coupling, and elicit the influence of extrinsic inputs on the states respectively. Here, for simplicity, we have expanded the state equation around $\vec{x} = 0$ and assumed that $f(0) = 0$.

The bilinear approximation of DCM is represented in terms of nonlinear differential equations while the GCM (see Eq. (1)) is formulated in discrete time and the dependencies among state variables are approximated by a linear mapping over time-lags which seems to be quite different. However, we can find the difference is that the bilinear form includes deterministic inputs and observation variables and equations which are not considered in GCM. The formulation (18) comes from the Volterra series and is certainly a more accurate and biophysical constraint representation of a biological system. On the other hand, the GCM with autoregressive representation always takes the past information into consideration while the bilinear approximation of DCM has no time-lags included in the differential equations although the general form of DCM may have [37]. So, if we alter the DCM to the form:

$$\begin{aligned} d\vec{x}(t) &= dt \int_0^t \{[A + u(t-\tau)B]\, \vec{x}(t-\tau) + u(t-\tau)\vec{c}\}\, k(\tau)d\tau + \sigma dB_t \\ \vec{y}(t) &= \tilde{g}(\vec{x}(t)) + \vec{\varepsilon}(t) \end{aligned} \tag{19}$$

where $k(\cdot)$ is a kernel function, then the DCM shares the feature of the GCM. On the other hand, our EGCM includes both deterministic inputs and observation variables thus takes

the advantages of both DCM and GCM. In the general form of EGCM (Eq. (2)), we can find that when $p = 1$, this is the discrete form of the bilinear form of DCM, and when $g(x) = x$, it is the GCM with additional inputs.

Advantage of extended approach. In contrast to all previous methods in estimating Granger causality in the literature where essentially a regression method is employed, in EGCM we incorporate noise observation variables and apply the extended Kalman filter to recover the state variables. Additional inputs are also included in EGCM on the basis of an autoregressive model. The advantage of such an approach over the previous methods is obvious. The EGCM is more reasonable when we are faced with experimental data affected by a particular stimulus and applicable to cases where we cannot track the state variables respectively but just a function of them, or where the observation noise is considerable. Comparing to the traditional VAR models, all the coefficients in EGCM correspond to intrinsic or latent dynamic coupling and changes induced by each input which endow the model with interpretability power. Furthermore, all the previous methods in estimating Granger causality are batch learning: they require collection of all data before an estimation can be made. The extended Kalman filter, on the other hand, is an online learning: we can now update Granger causality instantaneously. One may argue that this is a common feature of online learning vs. batch learning. However, it is novel in the context of Granger causality. When we are faced with biological data, this feature becomes particularly significant. As we know, adaptation, or learning in animals, is very important but this makes it difficult to analyze since adaptation introduces dynamic change into the system. The classical way of estimating Granger causality can cope with this difficulty by introducing sliding windows in analyzing data. Of course, to select an optimal window size is always

an issue in such an approach. However, in Kalman filtering, we can have the advantage of the connection of the state variables from the connection matrix and such an issue is automatically resolved.

In comparison with the bilinear approximation of DCM, the advantages of EGCM are the following: First, it allows time delay in the model more naturally and easier to deal with. Time delay is ubiquitous in a biological system, no matter whether we are considering gene, protein, metabolic and neuronal networks. Secondly, using Granger causality we are able to summarize the causal effect into a single number which is more transparent and easy to understand, particularly in a system with a time delay. Thirdly, it allows a frequency domain decomposition. We know that when we are dealing with a dynamic system it is sometimes much informative to view it in the frequency domain rather than in the time domain, as we have partly demonstrated here. Of course, since Eq. (19) is a continuous time version of Eq. (2), the results in the frequency domain obtained for Eq. (2) is essentially for the DCM model as well. We summarize our comparisons in Table 1 (see [10]).

Table 1 around here

Other types of data. In the current paper, we have only applied EGCM to LFP data although it is clearly applicable to many other types of biological data. For example, in gene microarray data, we can have a readout of transcriptional changes in several thousand genes at different times over a period of many hours [38]. The same is the case with multiple protein measurements over time in biological systems or in metabolic changes. In all these situations estimation of altered causal connections in both time and frequency domains will provide invaluable information about changes occurring in the relationships

between different components in the systems being studied.

IT Hemispheric Differences and Learning. The results of the EGCM analysis of our IT LFP data provide the first evidence for connectivity changes between and within left and right ITs as a result of face recognition learning. It is clear that learning is a dynamic and complex process [39,40]. In both sheep during learning there were more causal connections from the left to the right IT than vice-versa during learning trials. However, immediately after learning had occurred the number of left to right connections diminished to the same low level as seen from right to left. Within the hemispheres connectivity increased progressively over time in the right IT after learning but remained the same or decreased in the left IT. There was a strong negative correlation between the number of connections from left to right and the number within the right IT. This suggests that the left to right IT connections may exert some form of inhibitory control over the number within the right IT and that this therefore needs to be weakened for new face discriminations to be learned. The EGCM frequency analysis using theta and gamma oscillation data in the two hemispheres showed that during learning of new face pairs there was significantly more information being processed in the low frequency (theta) in the right IT than in the left. After learning however this declined and it appeared that the higher frequency information (gamma) became more dominant in both hemispheres. Lower frequency oscillations are more associated with global encoding over widespread areas of brain whereas higher frequencies are more associated with more localized encoding. This may suggest that during the course of learning new faces the right IT uses a more global mode of encoding to promote more rapid learning and that once learning has successfully occurred the right IT shifts to a more localized encoding strategy for maintaining learning. In the left IT on the other hand this

31

more local encoding strategy predominates both during and after learning. This is in broad agreement with recent proposals that the left hemisphere is more involved in local encoding and the right in global encoding [18, 19] although in the case of face recognition it would appear that the right hemisphere shifts from a global to local encoding strategy once faces have been learned. Clearly more analyses of this kind are required before these differences in left and right brain hemispheres processing and interactions can be fully understood but combining multiple LFP recordings and EGCM will be a powerful future approach.

Field-type Model. In the current paper, we have not explicitly introduced the spatio-correlation between each variables (electrodes). In other words, we have ignored the geometric relationship of electrodes in the array. This is certainly an over-simplification of the real situation due to the following reasons. First of all, despite the long history of multi-electrode array recordings, *in vivo* recording in, for example, IT is still very rare and difficult. Even we have a reliable recording session, the obtained data set is hard to fully analyze: for example, to reliably sort the spikes [41]. Secondly, assuming we could work out the spatio-temporal model for one animal, it is almost no sense to map the results about the detailed geometrical relationship (electrodes) to another animal. Also, in our experiments, we often face the situation that we have to discard the data from quite a few electrodes. A spatio-temporal (random field) approach as developed in fMRI [42] is not considered here. However, with the improvement of experimental techniques and the introduction of new techniques (see for example, functional multineuron calcium imaging [43]), a spatio-temporal model (a random field or a random point field) is cried for and is one of our future research topics. The well developed approach: dynamic expectation maximization [44], could play a vital role here.

References

[1] Cantone I, Marucci L, Iorio F, Ricci M, Belcastro V, et al. (2009) A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. Cell 137: 172–181.

[2] Camacho D, Collins J (2009) Systems biology strikes gold. Cell 137: 24–26.

[3] Zou C, Kendrick KM, Feng J (2009) The fourth way: Granger causality is better than the three other reverse-engineering approaches. Cell http://www.cell.com/comments/S0092-8674(09)00156-1.

[4] Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M (2008) Top-Down Control of Human Visual Cortex by Frontal and Parietal Cortex in Anticipatory Visual Spatial Attention. J Neurosci 28: 10056-10061.

[5] Seth A (2008) Causal networks in simulated neural systems. Cogn Neurodyn 2: 49-64.

[6] Seth AK, Edelman GM (2007) Distinguishing causal interactions in neural populations. Neural Computation 19: 910-933.

[7] Simpson EH (1951) The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society Series B (Methodological) 13: 238–241.

[8] Pearl J (1998) Causality: Models, Reasoning, and Inference. Cambridge, UK: Cambridge University Press.

[9] Zou C, Feng J (2009) Granger causality vs. dynamic bayesian network inference: a comparative study. BMC Bioinformatics 10: 122.

[10] Friston K (2009) Causal modelling and brain connectivity in functional magnetic resonance imaging. PLoS Biol 7: e1000033.

[11] David O, Guillemain I, Saillet S, Reyt S, Deransart C, et al. (2008) Identifying neural drivers with functional mri: An electrophysiological validation. PLoS Biol 6: e315.

[12] Friston K, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19: 1273 - 1302.

[13] Kendrick K, Zhan Y, Fischer H, Nicol AU, Zhang X, et al. (2009) Learning alters theta-nested gamma oscillations in inferotemporal cortex. Nature Precedings : hdl: 10101/ npre. 2009.3151.1.

[14] Peirce JW, Kendrick KM (2002) Functional asymmetry in sheep temporal cortex. NeuroReport 13: 2395–2399.

[15] Kendrick KM (2006) Brain asymmetries for face recognition and emotion control in sheep. Cortex 42: 96 - 98.

[16] Tate AJ, Fischer H, Leigh AE, Kendrick KM (2006) Behavioural and neurophysiological evidence for face identity and face emotion processing in animals. Philos Trans R Soc Lond B Biol Sci 361: 2155–2172.

[17] Shinohara Y, Hirase H, Watanabe M, Itakura M, Takahashi M, et al. (2008) Left-right asymmetry of the hippocampal synapses with differential subunit allocation of glutamate receptors. Proceedings of the National Academy of Sciences 105: 19498–19503.

[18] MacNeilage PF, Rogers LJ, Vallortigara G (2009) Evolutionary origins of your right and left brain. Scientific American Magazine : 6/24/09.

[19] Turgeon M (1994) Right-brain left-brain reflexology: a self-help approach to balancing life energies with color, sound, and pressure point techniques. Rochester,VT: Healing Arts Press.

[20] Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37: 414-428.

[21] Sun X, Jin L, Xiong M (2008) Extended kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks. PLoS ONE 3: e3758.

[22] Li P, Goodall R, Kadirkamanathan V (2004) Estimation of parameters in a linear state space model using a rao-blackwellised particle filter. Control Theory and Applications, IEE Proceedings 151: 727–738.

[23] Guo S, Wu J, Ding M, Feng J (2008) Uncovering interactions in the frequency domain. PLoS Comput Biol 4: e1000087.

[24] Schelter B, Winterhalder M, Timmer J (2006) Handbook of time series analysis: recent theoretical developments and applications. Weinheim: Wiley-VCH.

[25] Geweke J (1982) Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association 77: 304–313.

[26] Geweke J (1984) Measures of conditional linear dependence and feedback between time series. Journal of the American Statistical Association 79: 907–915.

[27] Wu J, Sinfield JL, Feng J (2009) Analysis of biological rhythms using a harmonic causal method. BMC Systems Biology (submitted) .

[28] Guo S, Seth AK, Kendrick KM, Zhou C, Feng J (2008) Partial granger causality–eliminating exogenous inputs and latent variables. Journal of Neuroscience Methods 172: 79–93.

[29] Chen Y, Rangarajan G, Feng J, Ding M (2004) Analyzing multiple nonlinear time series with extended granger causality. Physics Letters A 324: 26 - 35.

[30] Ladroue C, Guo S, Kendrick K, Feng J (2009) Beyond element-wise interactions: Identifying complex interactions in biological processes. PLoS ONE 4: e6899+.

[31] Wu J, Liu X, Feng J (2008) Detecting causality between different frequencies. Journal of Neuroscience Methods 167: 367 - 375.

[32] Gourvitch B, Bouquin-Jeanns RL, Faucon G (2006) Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. Biol Cybern 95: 349–369.

[33] David O, Guillemain I, Saillet S, Reyt S, Deransart C, et al. (2008) Identifying neural drivers with functional mri: An electrophysiological validation. PLoS Biol 6: e315.

[34] Arulampalam S, Maskell S, Gordon N (2002) A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing 50: 174–188.

[35] Nelson A (2007) Nonlinear estimation and modeling of noisy time-series by dual Kalman filtering methods. Ph.D. thesis.

[36] Buzsaki G (2006) Rhythms of the Brain. USA: Oxford University Press.

[37] David O, Kiebel S, Harrison L, Mattout J, Kilner J, et al. (2006) Dynamic causal modelling of evoked responses in EEG and MEG. NeuroImage 30: 1255–1272.

[38] Feng J, Yi D, Krishna R, Guo S, Buchanan-Wollaston V (2009) Listen to genes: Dealing with microarray data in the frequency domain. PLoS ONE 4: e5098+.

[39] Sedwick C (2009) Practice makes perfect: Learning mind control of prosthetics. PLoS Biol 7: e1000152.

[40] Robinson R (2009) From child to young adult, the brain changes its connections. PLoS Biol 7: e1000158.

[41] Horton PM, Nicol AU, Kendrick KM, Feng JF (2007) Spike sorting based upon machine learning algorithms (soma). Journal of Neuroscience Methods 160: 52–68.

[42] Valdes-Sosa PA (2004) Spatio-temporal autoregressive models defined over brain manifolds. Neuroinformatics 2: 239–250.

[43] Namiki S, Matsuki N, Ikegaya Y (2009) Large-scale imaging of brain network activity from $> 10,000$ neocortical cells. Nature Precedings : 2009.2893.1.

[44] Friston KJ, Trujillo-Barreto N, Daunizeau J (2008) Dem: A variational treatment of dynamic systems. NeuroImage 41: 849–885.

Fig. 1. Results on Toy Model 1. A. Traces of the time series. **B**. The causal relationships considered in Toy Model 1 between the three state variables. **C**. The estimated parameters $\tilde{a}_{11}^1, a_{22}^2$, and c_3 for the simulated data in Toy Model 1. The initial values of the three parameters are all set to 0. The covariance matrix $Z(t)$ is first set to decay slowly to achieve faster convergence and then set to decay faster after two hundred time points to ensure a better accuracy. **D**. Frequency decomposition of all kinds of relationships between the state variables. Significant causal influences are marked by red.

Fig. 2. Results on Toy model 2. Network structures with and without stimulus. **A**. Confidence intervals of all links between units. The data is generated with Eq. (15), but we use $p = 0$ (without input) in our algorithms and a traditional VAR(10) model to detect the causal influence. **B**. The network structure of the state variables corresponding to **A**. Two additional causal relationships are marked by the dashed line. **C**. Confidence intervals of all links between units. The data is generated with Eq. (16) where p and c_i, $i = 1, 2, 3$ are generated with normal distribution (with input). **D**. The network structure of the state variables corresponding to **C**.

Fig. 3. An example of the application of EGCM. The network detected by EGCM (top-panel) and the corresponding frequency decomposition (bottom-panel) for six randomly selected electrodes. In the frequency decomposition, significant causal influences are marked by red.

Fig. 4. Asymmetry between left and right hemisphere in the time domain. A. A

summary of the results in **B**, but locations in inferotemporal cortex are not precise, only for illustrative purposes. **B**. The mean connections from left hemisphere to right hemisphere, right hemisphere to left hemisphere and within both regions with the three bars corresponding to the results before learning (blue bar), after learning (green bar), and one month after learning (purple bar) in Sheep B. Significant changes after t-test are marked by arrows (right to left, all pairs are not significant, as indicated by "none"; within the right hemisphere, all pairs are significant, marked by "all") . For Sheep C, an additional bar (one week after learning) is added (the third bar). Only significant changes from left to right and within the right hemisphere are indicated by arrows. **C**. Statistic summaries of results in **B**.

Fig. 5. Asymmetry in the frequency domain interactions. A. Mean and maximum ratio using all the three sheep before and after learning. **B**. Upper panel: Mean and maximum ratio of sheep B (see Experiment subsection in Methods section) before learning, after learning and one month after learning (see Fig. 4). Bottom panel: Mean and maximum ratio of sheep C before learning (the first bar), immediately after learning (the second bar), one week after learning and one month after learning (the third and the fourth bar). **C**. Summaries of results in **B**.
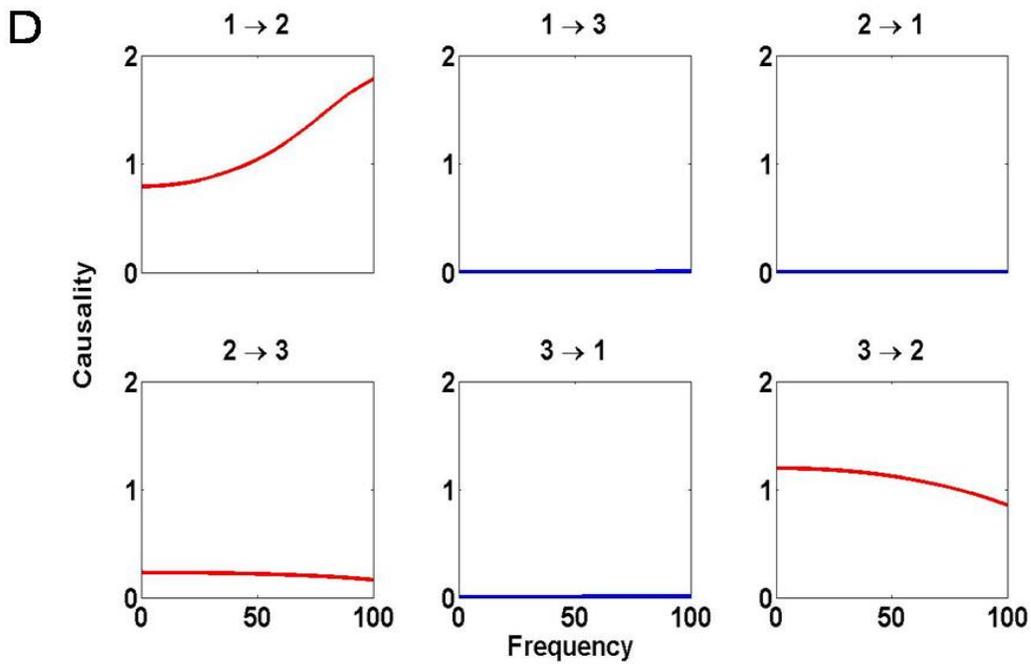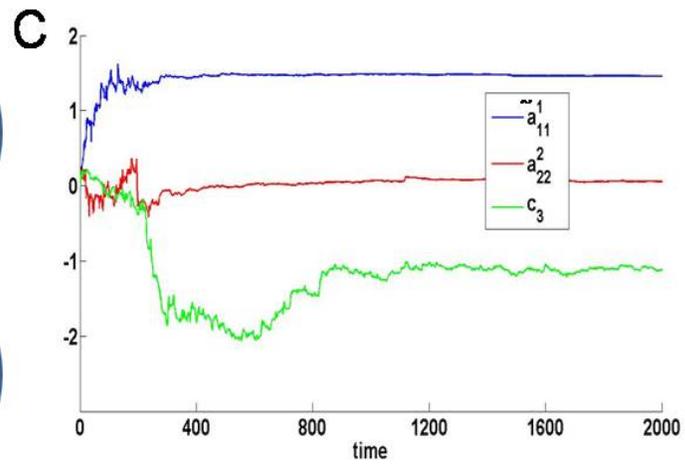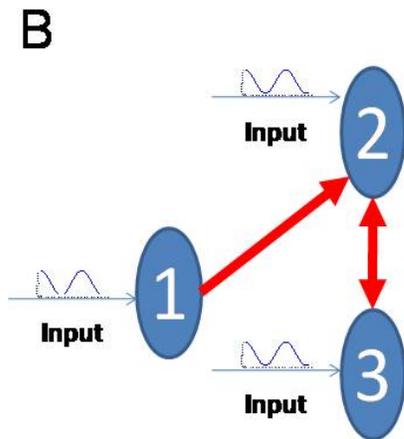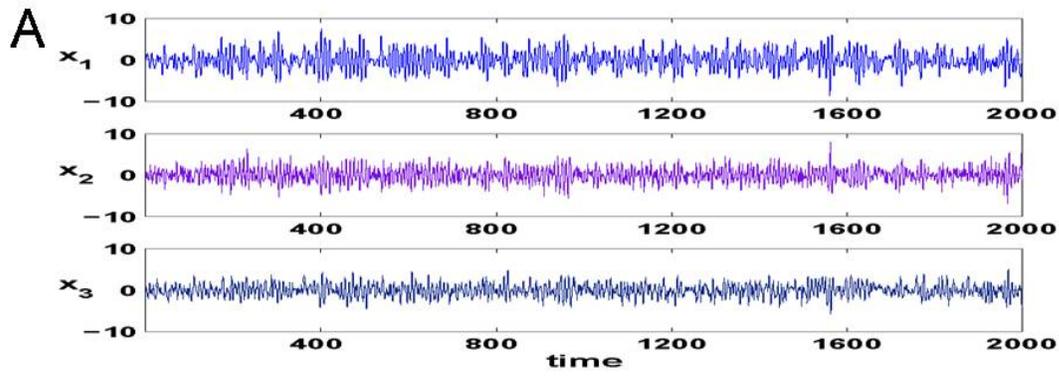
Table 1. Comparing DCM, GCM and EGCM.
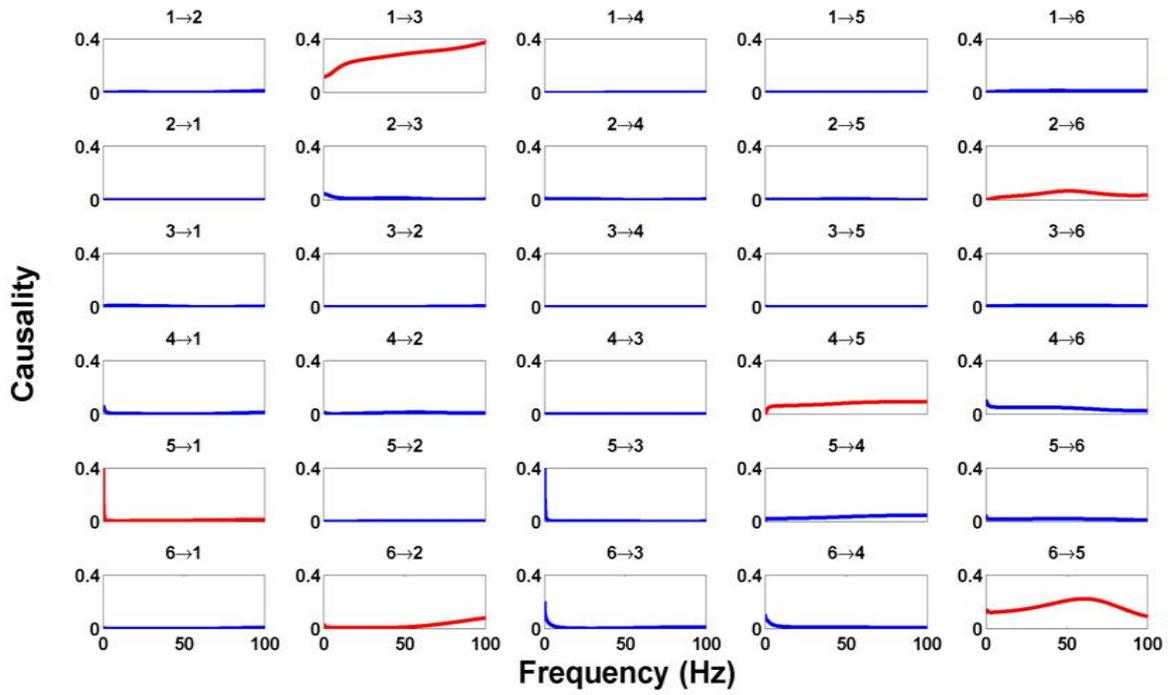
Figure 1:

41
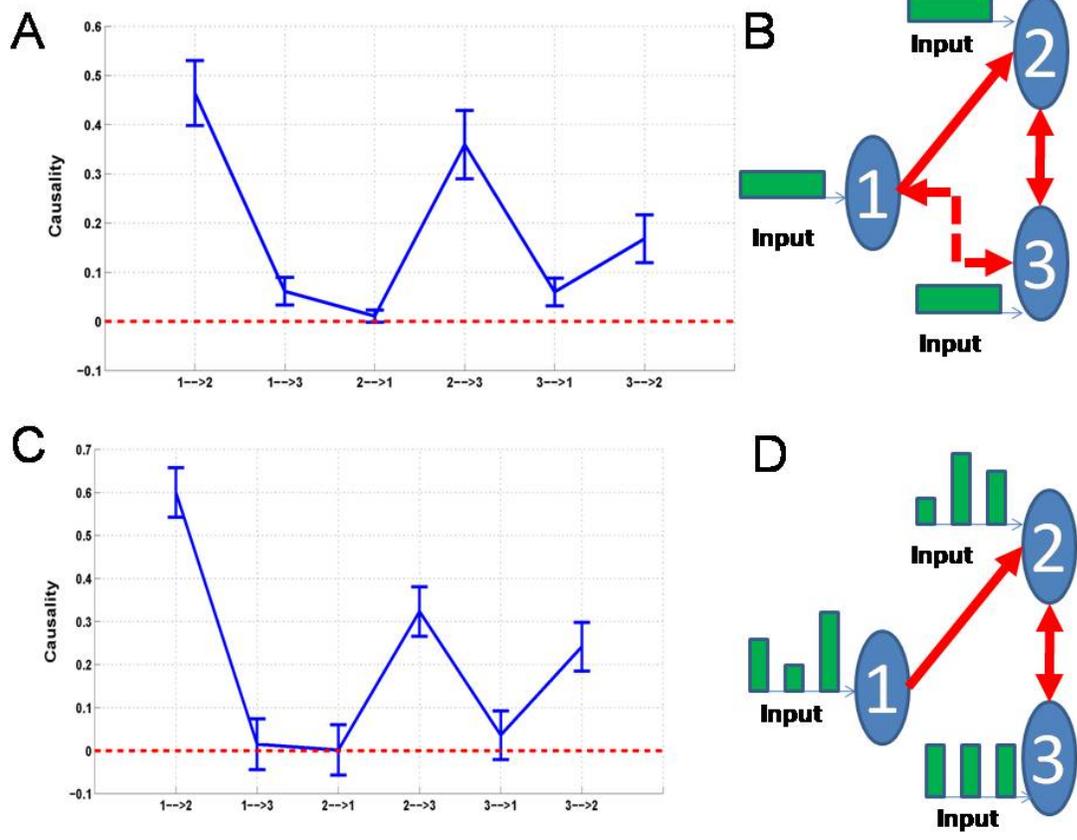
Figure 2:
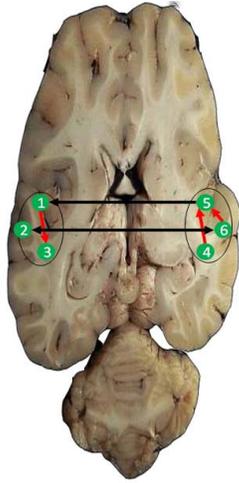
Figure 3:
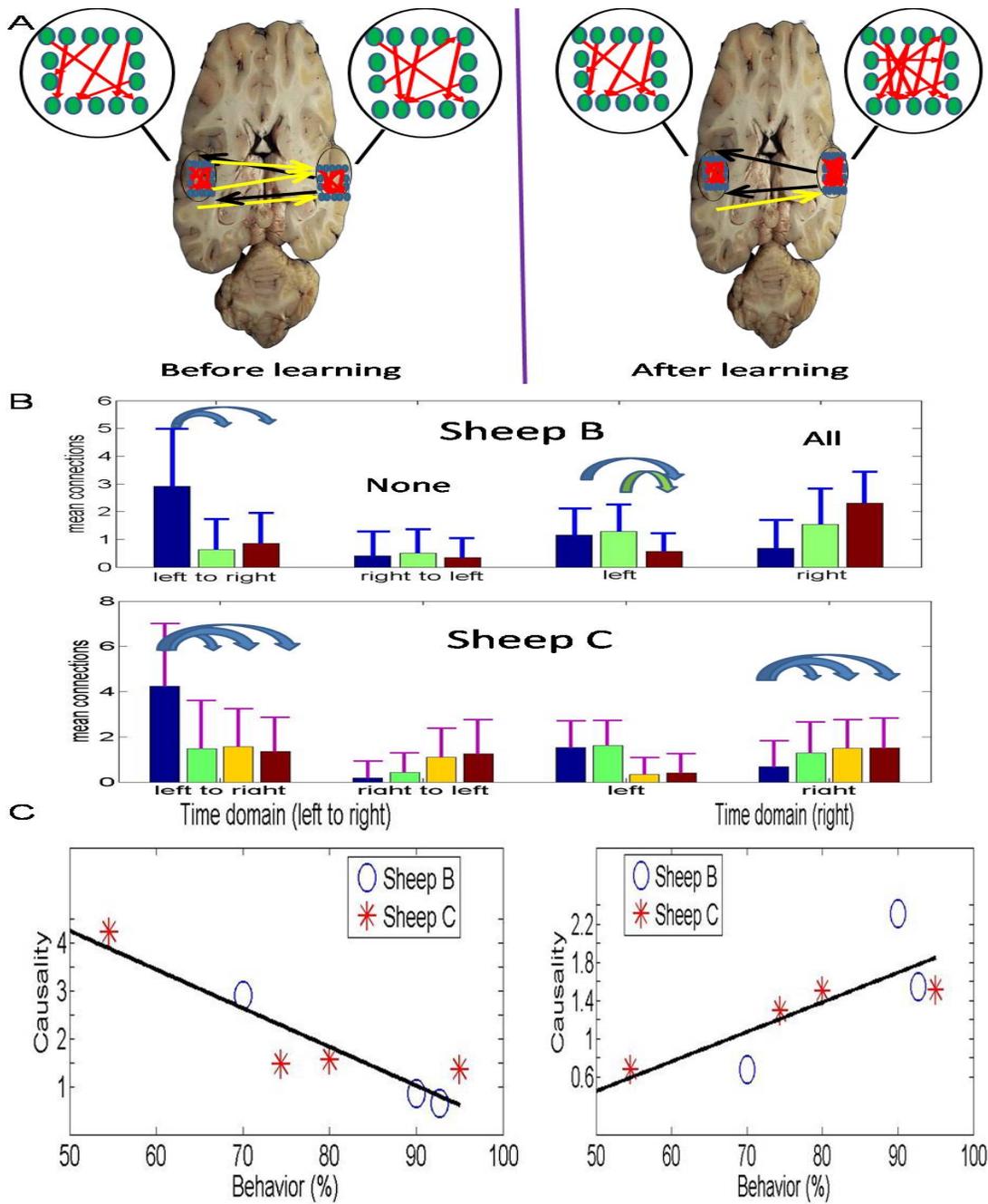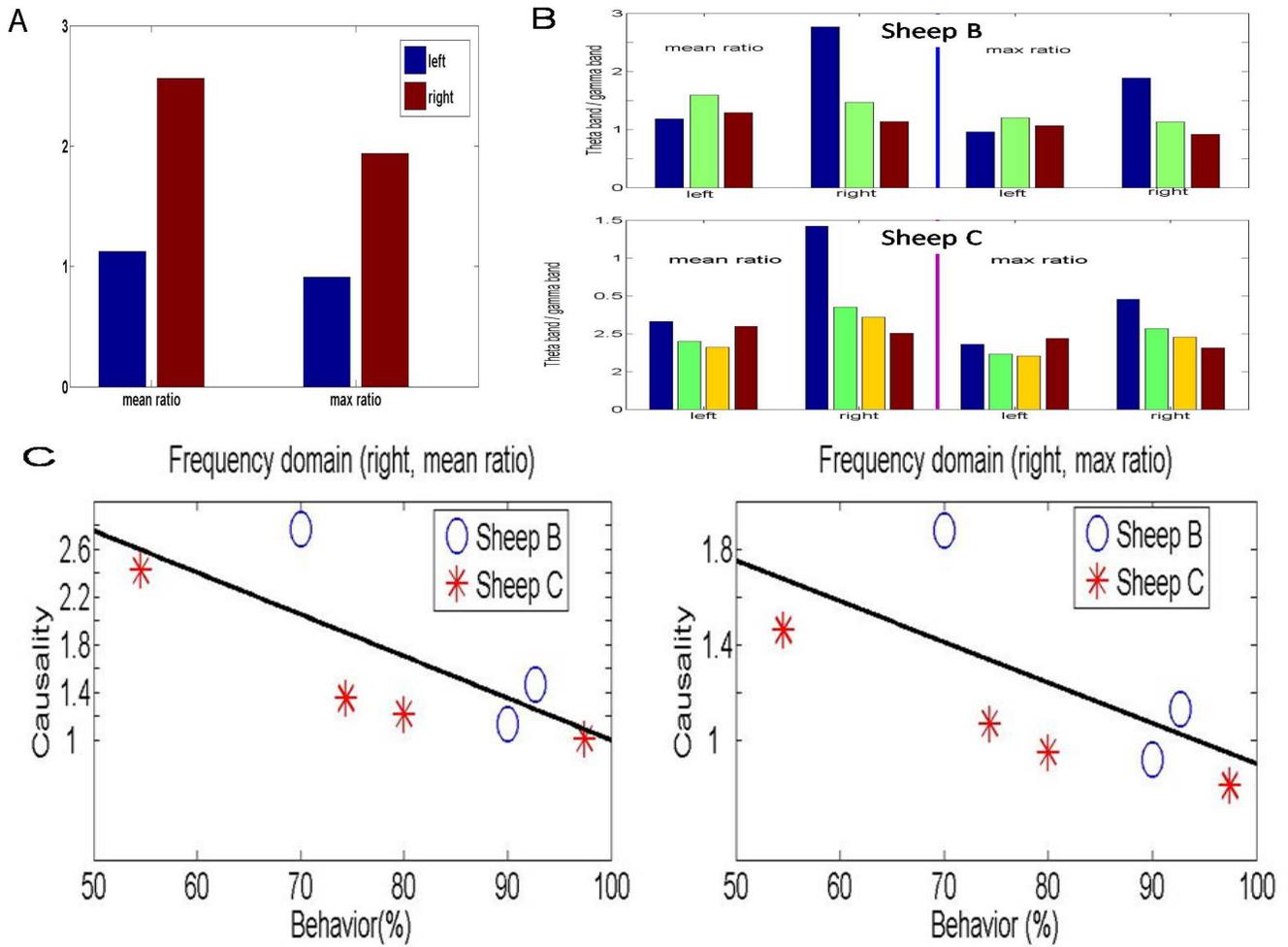
Figure 4:

Figure 5:

Commonalities and Differences between DCM, GCM and EGCM	DCM	GCM	EGCM
commonalities			
Multivariate analysis of time-series data	Yes	Yes	Yes
Models directed coupling	Yes	Yes	Yes
Inference on models	Yes	Yes	Yes
Frequency decomposition	Yes	Yes	Yes
Differences			
Causality based on temporal precedence	No	Yes	EGCM is more general
Causality based on control theory	Yes	No	Yes
Requires known inputs	Yes	No	In general yes, but see example 2
Requires orthogonal innovations	No	Yes	Not necessary
Requires stationary processes	No	Yes	Could use sliding window
Requires a specific biophysical model	Yes	No	Yes
Models non-linear coupling	Yes	No	Yes
Inference on model parameters	Yes	No	Yes